ASYMPTOTIC INFERENCE IN A NONLINEAR COINTEGRATED MODEL

PJeganathan

Indian Statistical Institute

Bangalore Center

8th Mile, Mysore Road

Bangalore 560059

India

Preliminary and incomplete.

Prepared for a seminar at the Department of Economics, Yale University.

1 INTRODUCTION AND MOTIVATION

This talk consists of some extensions and generalizations of some of the results in Park and Phillips ("Nonlinear regression with integrated processes.", *Econometrica*, **69**, 2001).

A familiar and a simplest formulation of the cointegrating model assumes the availability of two economic or financial variables X_k and $Y_k, k = 1, \ldots, n$, such that

$$Y_k - \beta X_k = \text{stationary} \tag{1}$$

where X_k is nonstationary, usually an integrated process. The relationship (1) means that X_k and Y_k are linearly related except for the unknown parameter β. We shall call X_k an exogenous variable.

To indicate one possible limitation of this formulation, it may be recalled that economic or financial variables are social phenomena, and are governed to a large extent by the psychology of (the members that constitute the) social institutions. Even though it may appear at a very broad level that the relationship between two or more variables is linear, or similar mathematically ideal relationship, it is also usual to see that one variable tend to react in a manner different from that of the general trend when certain other variable becomes close to a certain threshold. In fact one might argue that the overall trend itself is governed by such reactions (by the independently acting members of the social institution to a fairly large number of variables in a complex manner); one gets the appearance of a trend over certain regions when the effect of the heterogeneity of the nature of the (independent) reactions is smoothed or averaged out (by the usual law of large numbers or mass scale phenomenon). But the effect might be distinctly pronounced (due to a conscious or deliberate reaction) when a certain exogenous variable becomes close to a threshold, to the extent that the overall trend cannot capture such effects. It is also important to note that the moment of the approach of the exogenous variable to the threshold itself is completely random.

Now assume for simplicity that the threshold is the origin 0. Then the region of the indicated reaction will be of the form $\mathbb{I}\left(-a_n^{-1}c < X_k < a_n^{-1}d\right)$ for some positive constants c and d (where $\mathbb{I}(.)$ stands for an indicator function). Here the constant a_n may be viewed as a sort of bandwidth. (It will be assumed below that for a suitable $0 < H < 1$, the process $n^{-H}X_{[nt]}$ converges in distribution; in particular the variability of X_k increases with k. The a_n will be such that $a_n n^H \to \infty$ but $a_n n^{H-1} \to 0$. Thus a_n can be allowed to remain constant, or $\to 0$ or $\to \infty$ appropriately.) Thus the effect of this reaction may be expressed in the form of a function $f(\gamma, a_n X_k)$ where γ is an unknown parameter and $f(\gamma, x)$will in effect have a compact support with respect to the variable x. Therefore, in place of the relationship (1), the following more general form

$$Y_k - \beta X_{k-1} - f(\gamma, a_n X_{k-1}) = \text{ stationary}$$

will be considered.

Still more generally, in place of $\mathbb{I}\left(-a_n^{-1}c < X_k < a_n^{-1}d\right)$, one may consider the region of the form $\mathbb{I}\left(-a_n^{-1}c\left(\frac{k}{n}\right) < X_k + n^H \ell\left(\frac{k}{n}\right) < a_n^{-1}d\left(\frac{k}{n}\right)\right)$ where $\ell(t)$ represents an appropriate moving threshold and similarly $c(t)$ and $c(t)$ form appropriately curved boundaries. It may also be necessary to consider more than one such regions. We shall not deal with such generalizations here.

It may also be noted that the relationship

$$Y_k - \beta X_k \mathbb{I}\left(-a_n^{-1}c < X_k < a_n^{-1}d\right) = \text{ stationary}$$

may be viewed to have approximately the same role of the relationship $Y_k - \beta X_k =$ stationary, when a_n is allowed to tend to zero appropriately. (For instance choose $a_n = n^{-H}\log n$.)

Another interesting point that is likely to have some significance in modeling is that $\sum_{k=1}^{n}\mathbb{I}\left(-a_{n1}^{-1}c < X_k < a_{n1}^{-1}d\right)$ will be asymptotically independent of $\sum_{k=1}^{n}\mathbb{I}\left(-a_{n2}^{-1}c < X_k < a_{n2}^{-1}d\right)$ whenever $\frac{a_{n1}}{a_{n2}} \to 0$. By this we mean, keeping in mind the remark of the preceding paragraph, a relationship of the form

$$Y_k - f_1(\gamma_1, a_{n1}X_{k-1}) - \ldots . f_l(\gamma_l, a_{nl}X_{k-1}) = \text{ stationary}$$

for suitable functions $f_1(\gamma_1, x), \ldots\ f_l(\gamma_l, x)$, where $\gamma_1, \ldots, \gamma_l$ are the parameters and $a_{n1} < \ldots < a_{nl}$ are such that $\frac{a_{n,i-1}}{a_{n,i}} \to 0, i = 2, \ldots l$, would be more realistic.

In Section 2, a precise form of the model will be introduced. In the same section the limiting behavior of certain random quantities that will be involved in obtaining the asymptotic behavior of the estimators will be recalled (from the first lecture). Two methods of construction of estimators of the parameters will be dealt with. Section 3 deals with the first method, which is the same as the one dealt with in Park and Phillips (2001). Section 4 deals with the second method; it gives the approximate MLE as a particular case. Both methods have certain advantages as well as certain disadvantages.Section 5 shows that the likelihood has an LAMN approximating structure.

1 STATEMENTS OF THE RESULTS

Consider two-dimensional observations $(X_k, Y_k), k = 1, \ldots, n$, such that

$$Y_k = \beta X_{k-1} + f(\gamma, a_n X_{k-1}) + u_k$$

with the constants a_n are as motivated in Section 1 (and satisfying the condition (9) below), and

$$X_k = \rho X_{k-1} + v_k.$$

Here β, γ and ρ are unknown parameters. The complete specification of the conditions on the function f will be given later, but it will include that $f(\gamma, x)$ is differentiable with respect to γ and, denoting the resulting derivative by $f_{(1)}(\gamma, x)$, the function $x \longmapsto f_{(1)}(\gamma, x)$ is locally Riemann integrable such that

$$\left|f_{(1)}(\gamma, x)\right| \leq \frac{K}{1 + |x|^{1+c}}, \text{ for some } c > 0 \text{ as } |x| \to \infty.$$

The errors u_k and v_k are assumed to have the following restrictions. The $u_k, k \geq 1$, are i.i.d. with $E\left[u_1\right] = 0$ and $E\left[u_1^2\right] = 1$, and $v_k, -\infty < k < \infty$, form a linear process of the form

$$v_k = \sum_{j=0}^{\infty} c_j(\alpha) \xi_{k-j} \tag{2}$$

where ξ_j, $-\infty < j < \infty$ are i.i.d such that $(u_k, \xi_k), k \geq 1$, are i.i.d, independent of $(\xi_j, -\infty < j \leq 0)$ with $E[\xi_1] = 0$ and $E[\xi_1^2] = 1$. Here α is an unknown parameter but will be assumed to be known for simplicity. In addition v_k are invertible so that

$$\xi_k = \sum_{j=0}^{\infty} d_j(\alpha) v_{k-j}. \tag{3}$$

The inference procedures will be constructed under the truncated (misspecified) model assumption

$$v_k = \sum_{j=0}^{k-1} c_j(\alpha) \xi_{k-j} \tag{4}$$

and

$$\xi_k = \sum_{j=0}^{k-1} d_j(\alpha) v_{k-j}. \tag{5}$$

The nature of the resulting misspecification will be indicated later.

We shall distinguish three cases:

Case I. In this case, it is assumed that

$$\sum_{j=0}^{\infty} |c_j(\alpha)| < \infty \text{ and } \sum_{j=0}^{\infty} |d_j(\alpha)| < \infty$$

such that

$$\left| \sum_{j=0}^{\infty} c_j(\alpha) \right| > 0.$$

This is the situation considered in Park and Phillips (2001).

Case II. In this case v_k are assumed to follow a fractional ARIMA$(0, \alpha, 0)$, that is,

$$(1 - B)^{\alpha} v_k = \xi_k$$

with $-1/2 < \alpha < 1/2$, $\alpha \neq 0$ where B is the back-shift operator $Bv_k = v_{k-1}$. This means (2) and (3) hold with

$$c_j(\alpha) = (\alpha)(1 + \alpha) \cdots (j - 1 + \alpha)/j! \ \sim \ \frac{1}{(\alpha - 1)! j^{1-\alpha}} \text{ as } j \to \infty \tag{6}$$

and

$$d_j(\alpha) = c_j(-\alpha) \; \sim \; \frac{1}{(-\alpha - 1)!j^{1+\alpha}} \text{ as } j \to \infty. \tag{7}$$

It is important to note that

$$\sum_{j=0}^{\infty} c_j(\alpha) = 0 \text{ when } \; -1/2 < \alpha < 0.$$

It is possible to consider the more general situation in which $c_j(\alpha) = K(\alpha) j^{\alpha-1} R(j)$ for some slowly varying function $R(j)$ (and $\sum_{j=0}^{\infty} c_j(\alpha) = 0$ when $-1/2 < \alpha < 0$). This will include the fractional ARIMA(p, α, q) case, in which case $R(j) \sim 1$ as $j \to \infty$.

Case III: In this case the model specified by (4) and (5) with $c_j(\alpha)$ as in (6) and with $d_j(\alpha)$ as in (7) is actually taken as the true model.

For simplicity of presentation we shall assume throughout that $\rho = 1$. In addition we shall assume for simplicity that the parameter α is known. More generally, one can take ρ to be in an unknown vicinity of 1 and can be simultaneously estimated with the parameters β and γ but it is important to note however that the limiting behaviors of the estimators of β and γ will not remain the same as those when ρ is assumed to be known $\rho = 1$. The parameter α can also be estimated simultaneously, and whether α is assumed to be known or estimated, the asymptotic behavior of the estimators of the other parameters will remain the same, and vice versa.

Then note in particular that

$$X_k = \sum_{j=1}^{k} v_j.$$

The following limiting behaviors follow from Lecture I. Define

$$H = \begin{cases} \frac{1}{2} & \text{in the Case I} \\ \alpha + \frac{1}{2} & \text{in the case II or III.} \end{cases}$$

Statement A: Let $\psi(x)$ be such that $E[\psi(u_1)] = 0$ and $E[\psi^2(u_1)] = \lambda > 0$. Then the finite dimensional distributions of the process

$$\left(n^{-\frac{1}{2}}\sum_{k=1}^{[nt]}\psi\left(u_k\right),\, n^{-H}X_{[nt]}^2\right)$$

will converge in distribution to those of

$$\left(W\left(t\right),\, B_H\left(t\right)\right)$$

where $W\left(t\right)$ is a Brownian motion with variance λ and $B_H\left(t\right)$ is a fractional Brownian motion (FBM) such that

$$\operatorname{Cov}\left(W\left(t\right), B_H\left(s\right)\right) = b\operatorname{Cov}\left(\psi\left(u_1\right), \xi_1\right)\int_0^{\min(s,t)}\left(s-u\right)^{H-1/2}du.$$

Here recall that

$$B_H\left(t\right) = \begin{cases} \left(\sum_{j=0}^{\infty}c_j(\alpha)\right)B\left(t\right) & \text{in the Case I} \\ b\int_{-\infty}^{0}\left[\left(t-u\right)^{H-1/2}-\left(-u\right)\right]dB\left(t\right)+b\int_0^t\left(t-u\right)^{H-1/2}dB\left(t\right) & \text{in the Case II} \\ b\int_0^t\left(t-u\right)^{H-1/2}dB\left(t\right) & \text{in the Case III} \end{cases}$$

where the Brownian motion $B\left(t\right),\ -\infty < t < \infty$, is the limit of the process $n^{-\frac{1}{2}}\sum_{k=1}^{[nt]}\xi_k$. (We make the convention that in the Case I, $b = \sum_{j=0}^{\infty}c_j(\alpha)$.)

This will in particular entail that

$$\left(n^{-H-\frac{1}{2}}\sum_{k=1}^{n}X_{k-1}\psi\left(u_k\right),\, n^{-2H-1}\sum_{k=1}^{n}X_{k-1}^2\right)$$

converge in distribution to

$$\left(\int_0^1 B_H\left(t\right)dW\left(t\right),\, \int_0^1 B_H^2\left(t\right)dt\right).$$

Statement B: Let $g\left(x\right)$ be such that $x\longmapsto g^2\left(x\right)$ is locally Riemann integrable and $x\longmapsto \sup\left\{g^2\left(y\right): \left|x-y\right|\leq\eta\right\}$ is Lebesgue integrable for some $\eta>0$. Assume further that the Cramer's condition $\limsup_{|s|\to\infty}\left|E\left[e^{is\xi_1}\right]\right|<1$ is satisfied. Define

$$\delta_n = \sqrt{a_n n^{H-1}}. \tag{8}$$

Assume that

$$\delta_n \to 0 \text{ and } a_n n^H \to \infty. \tag{9}$$

Then

$$\left(\delta_n \sum_{k=1}^{n} g(a_n X_{k-1})\psi\left(u_k\right),\ \delta_n^2 \sum_{k=1}^{n} g^2(a_n X_{k-1})\right)$$

converges in distribution to

$$\left(Z\sqrt{\lambda L\left(1,0\right)\int_{-\infty}^{\infty} g^2(x)dx},\ \lambda L\left(1,0\right)\int_{-\infty}^{\infty} g^2(x)dx\right)$$

where $L\left(t,0\right)$ is the *local time (in* 0*) process* of the process $B_H\left(t\right)$, and Z is a standard normal r.v. *independent* of $L\left(t,0\right)$.

Recall that the local time $L(t,x)$ is defined to be such that

$$\lim_{\eta \downarrow 0} \frac{1}{\eta}\int_0^t \mathbb{I}_{[x,x+\eta)}(\Lambda_{\alpha,H}(s))ds = L(t,x) \qquad \text{in } \mathbf{L}^2 \text{ or a.s.}$$

for all t and x, see the notes for Lecture I for the detailed motivation and further details.

3 CONSTRUCTION OF ESTIMATORS: METHOD I

Two methods of construction of the estimators will be considered. The first method, related to the procedure considered in Park and Phillips (2001),will include the *least squares* (LS) estimators defined by

$$\left(\widetilde{\beta}_n, \widetilde{\gamma}_n\right) = \operatorname{argmin}_{(\beta,\gamma)}\left(Y_k - \beta X_{k-1} - f(\gamma, a_n X_{k-1})\right)^2.$$

To see the asymptotic nature of $\left(\widetilde{\beta}_n, \widetilde{\gamma}_n\right)$, assume for the moment that

$$f(\gamma, a_n X_{k-1}) = \gamma g(a_n X_{k-1}). \tag{10}$$

Then

$$\begin{pmatrix}\widetilde{\beta}_n \\ \widetilde{\gamma}_n\end{pmatrix} = \left(\sum_{k=1}^{n}\begin{pmatrix} X_{k-1}^2 & X_{k-1}g(a_n X_{k-1}) \\ X_{k-1}g(a_n X_{k-1}) & g^2(a_n X_{k-1})\end{pmatrix}^{-1}\right)\sum_{k=1}^{n}\begin{pmatrix} X_{k-1}Y_k \\ g(a_n X_{k-1})Y_k\end{pmatrix}$$

so that

$$\begin{pmatrix}\widetilde{\beta}_n-\beta\\ \widetilde{\gamma}_n-\gamma\end{pmatrix}=\left(\sum_{k=1}^{n}\begin{pmatrix}X_{k-1}^2 & X_{k-1}g(a_nX_{k-1})\\ X_{k-1}g(a_nX_{k-1}) & g^2(a_nX_{k-1})\end{pmatrix}^{-1}\right)\sum_{k=1}^{n}\begin{pmatrix}X_{k-1}u_k\\ g(a_nX_{k-1})u_k\end{pmatrix}$$

where recall that $u_k=(Y_k-\beta X_{k-1}-\gamma g(a_nX_{k-1}))$.

We now **claim** that

$$\delta_n n^{-H-\frac{1}{2}}\sum_{k=1}^{n}X_{k-1}g(a_nX_{k-1})=o\left(1\right). \tag{11}$$

Suppose for the moment that this claim is true. Then (recall from (8) that $\delta_n=\sqrt{a_nn^{H-1}}$) it follows from Statements A and B (with $\psi\left(x\right)=x$) of Section 2 that

$$\begin{pmatrix}n^{H+\frac{1}{2}}\left(\widetilde{\beta}_n-\beta\right)\\ \delta_n^{-1}\left(\widetilde{\gamma}_n-\gamma\right)\end{pmatrix}=\begin{pmatrix}\left(n^{-2H-1}\sum_{k=1}^{n}X_{k-1}^2\right)^{-1}n^{-H-\frac{1}{2}}\sum_{k=1}^{n}X_{k-1}u_k\\ \left(\delta_n^2\sum_{k=1}^{n}g^2(a_nX_{k-1})\right)^{-1}\delta_n\sum_{k=1}^{n}g(a_nX_{k-1})u_k\end{pmatrix}+o_p\left(1\right). \tag{12}$$

together with the fact that $\delta_n^{-1}\left(\widetilde{\gamma}_n-\gamma\right)$ will converge in distribution to

$$\left(L\left(1,0\right)\int_{-\infty}^{\infty}g^2(x)dx\right)^{-1/2}Z \tag{13}$$

which is a mixture of normal distribution, and $n^{H+\frac{1}{2}}\left(\widetilde{\beta}_n-\beta\right)$ will converge in distribution to

$$\left(\int_0^1 B_H^2\left(t\right)dt\right)^{-1}\int_0^1 B_H\left(t\right)dW\left(t\right) \tag{14}$$

which is in general not a mixture of the normal distribution; in fact as is clear from its representation that it has the usual ‘unit root’ limiting distribution.

REMARK: It is important to note that the representation (12) means the Wald quadratic statistic for testing β will have the same asymptotic behavior weather (the nuisance parameter) γ is assumed be known or unknown, that is, it can be constructed as if γ is known and then replace the unknown γ by its estimator. The same statement holds with regard to testing for γ (with β treated as a nuisance parameter). In the case of γ the asymptotic distribution of the Wald quadratic will be central Chi-squared under the null hypothesis. Unfortunately this is not the case with regard to the testing for β.

Now let us show that the claim (11) is true. Assume for the moment that $g(x)$ in (10) has a compact support for the argument x. Then $X_{k-1}g(a_nX_{k-1})$ is bounded in absolute value by $Ca_n^{-1}|g(a_nX_{k-1})|$ for some constant C. Also recall that $\delta_n^2\sum_{k=1}^n|g(a_nX_{k-1})|$ converges in distribution, using the Statement B of Section 2. Thus

$$\begin{aligned}
&\left|\delta_n n^{-H-\frac{1}{2}}\sum_{k=1}^n X_{k-1}g(a_nX_{k-1})\right| \\
\leq\ & Ka_n^{-1}\delta_n^{-1}n^{-H-\frac{1}{2}}\left(\delta_n^2\sum_{k=1}^n|g(a_nX_{k-1})|\right) \\
=\ & O_p\left(a_n^{-1}\delta_n^{-1}n^{-H-\frac{1}{2}}\right).
\end{aligned}$$

Now, noting that $\delta_n = \sqrt{a_nn^{H-1}}$ one has $a_n^{-1}\delta_n^{-1}n^{-H-\frac{1}{2}} = \sqrt{\frac{n}{a_n^3n^{H+2H+1}}} \to 0$, because of the assumption $a_nn^H \to \infty$. Thus (11) is true. This remains true under the more general condition that

$$|g(x)| \leq \frac{K}{1+|x|^{1+c}},\ \text{for some } c>0 \text{ as } |x|\to\infty \tag{15}$$

in view of the fact that, when $L>0$ is sufficiently large

$$\begin{aligned}
&\delta_nn^{-H-\frac{1}{2}}\sum_{k=1}^n|X_{k-1}g(a_nX_{k-1})|\,\mathbb{I}\left(|a_nX_{k-1}|>L\right) \\
\leq\ & K\delta_nn^{-H-\frac{1}{2}}\sum_{k=1}^n|X_{k-1}|\left(1+|a_nX_{k-1}|^{1+c}\right)^{-1}\mathbb{I}\left(|a_nX_{k-1}|>L\right) \\
\leq\ & K\delta_nn^{-H-\frac{1}{2}}\sum_{k=1}^n a_n^{-1}L^{-c} \\
=\ & KL^{-c}\delta_nn^{-H-\frac{1}{2}}na_n^{-1} \\
=\ & KL^{-c}a_n^{1/2}n^{(H-1)/2}n^{-H-\frac{1}{2}}a_n^{-1}n \\
=\ & KL^{-c}a_n^{-1/2}n^{-H/2}\to 0
\end{aligned}$$

because of the assumption $a_nn^H\to\infty$. Thus the claim (11) is true under the assumption (15).

It can be shown that the representation (12), as well as its limiting behavior, remain true without the linear restriction (10), in which case $f_{(1)}(\gamma, a_nX_{k-1})$, the first derivative of $f(\gamma,x)$ with respect to γ, will replace the role of $g(a_nX_{k-1})$, that is, one has

$$\begin{pmatrix} n^{H+\frac{1}{2}}\left(\widetilde{\beta}_n-\beta\right) \\ \delta_n^{-1}\left(\widetilde{\gamma}_n-\gamma\right)\end{pmatrix} = \begin{pmatrix} \left(n^{-2H-1}\sum_{k=1}^n X_{k-1}^2\right)^{-1} n^{-H-\frac{1}{2}}\sum_{k=1}^n X_{k-1}u_k \\ \left(\delta_n^2\sum_{k=1}^n f_{(1)}^2(\gamma,a_nX_{k-1})\right)^{-1}\delta_n\sum_{k=1}^n f_{(1)}(\gamma,a_nX_{k-1})u_k \end{pmatrix} + o_p(1).$$

Once it is known that the LS estimators $\widetilde{\beta}_n$ and $\widetilde{\gamma}_n$ are well behaved, one can construct more general class of so called approximate M-estimators as follows. Let $\psi(x)$ be an "influence function" such that

$$E\left[\psi(u_1)\right]=0,\ 0<E\left[\psi^2(u_1)\right]<\infty \text{ and } 0<\kappa=\left.\frac{dE\left[\psi(u_1+z)\right]}{dz}\right|_{z=0}<\infty.$$

Define

$$\begin{pmatrix}\widehat{\beta}_n \\ \widehat{\gamma}_n\end{pmatrix} = \begin{pmatrix}\widetilde{\beta}_n \\ \widetilde{\gamma}_n\end{pmatrix} + \left(\kappa\sum_{k=1}^n \widetilde{W}_{nk}\widetilde{W}_{nk}^T\right)^{-1}\sum_{k=1}^n \widetilde{W}_{nk}\psi\left(Y_k-\widetilde{\beta}_nX_{k-1}-f(\widetilde{\gamma}_n,X_{k-1}\right)$$

where

$$\widetilde{W}_{nk}=\begin{pmatrix} X_{k-1} \\ f_{(1)}(\widetilde{\gamma}_n,a_nX_{k-1})\end{pmatrix}$$

This is just the usual one-step Newton iteration procedure, with $\left(\widetilde{\beta}_n,\widetilde{\gamma}_n\right)$ as the preliminary estimators. Then it can be shown that

$$\begin{pmatrix} n^{H+\frac{1}{2}}\left(\widehat{\beta}_n-\beta\right) \\ \delta_n^{-1}\left(\widehat{\gamma}_n-\gamma\right)\end{pmatrix} = \begin{pmatrix} \kappa^{-1}\left(n^{-2H-1}\sum_{k=1}^n X_{k-1}^2\right)^{-1} n^{-H-\frac{1}{2}}\sum_{k=1}^n X_{k-1}\psi(u_k) \\ \kappa^{-1}\left(\delta_n^2\sum_{k=1}^n f_{(1)}^2(\gamma,a_nX_{k-1})\right)^{-1}\delta_n\sum_{k=1}^n f_{(1)}(\gamma,a_nX_{k-1})\psi(u_k) \end{pmatrix} + o_p(1). \tag{16}$$

The limiting distribution of this will be as given in (13) and (14) (except for the constant κ^{-1}).

Note that in some situations the preceding M-estimator corresponding to the influence function $\psi(x)$ may also approximately be viewed as

$$\left(\widehat{\beta}_n,\widehat{\gamma}_n\right)=\text{argmin}_{(\beta,\gamma)}\varrho\left(Y_k-\beta X_{k-1}-f(\gamma,X_{k-1})\right)$$

where $\frac{d\varrho(x)}{dx}=\psi(x)$. When $\varrho(x)=|x|^2$ this reduces to the LS estimator.

It may further be noted that when the common distribution of the i.i.d. errors η_k is assumed to have a heavy tailed distribution, it may not be appropriate to start with the LS estimators. However, starting with for example a least absolute deviation estimators, or any other estimators corresponding to an influence function $\psi(x)$ with $\sup_x |\psi(x)| < \infty$, will work well.

4 CONSTRUCTION OF ESTIMATORS: METHOD II

We now describe the second method of construction of estimators. This is related to approximate likelihood based estimators, and gives approximate MLE as a particular case. The method is very satisfying in the Cases I and III, in particular the Wald quadratics for testing the respective parameters β and γ, will, in addition to satisfying the remark made in Section 3, have asymptotic central Chi-squared distributions under the respective null hypotheses. But unfortunately in the Case II the procedure has some disadvantages, as will be indicated later.

In order to introduce the method, let $\psi(x, y)$ be a function of two variables such that

$$E\left[\psi(\eta_1, \xi_1) | \xi_1\right] = 0 \text{ a.s}, \; 0 < \lambda = E\left[|\psi(\eta_1, \xi_1)|^2\right] < \infty \tag{17}$$

and

$$0 < \kappa = \left.\frac{dE\left[\psi(u_1 + z, \xi_1)\right]}{dz}\right|_{z=0} < \infty.$$

One can construct several examples of such $\psi(x, y)$ as follows.

To describe the first example, which will lead to the approximate MLE, let $p(x, y)$ be the joint (Lebesgue) density function of (η_1, ξ_1). Then we take

$$\psi(x, y) = -\frac{\partial \log p(x, y)}{\partial x}.$$

Note that $-\psi(x, y)$ is just the derivative of the conditional density function of η_1 given $\xi_1 = y$. Now impose the restriction $E\left[|\psi(\eta_1, \xi_1)|^2\right] < \infty$. It is known that this restriction entails that $E\left[\psi(\eta_1, \xi_1) | \xi_1\right] = 0$ a.s. Note that the M-estimator with $\psi(x, y)$ as the influence function is just an approximate MLE, see below.

To describe the next example, which will lead to LS estimators, we take the joint density $p(x,y)$ to be of the spherically symmetric form

$$p(x,y) = |\det \Omega|^{-1/2} h\left(\left|\Omega^{-1/2}\begin{pmatrix} x \\ y \end{pmatrix}\right|\right)$$

for some function h, where

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix},$$

assumed positive definite, is the variance-covariance matrix of (η_1, ξ_1), that is, ω_{11}, ω_{12} and ω_{22} respectively are $E\left[|\eta_1|^2\right]$, $E[\eta_1\xi_1]$ and $E\left[|\xi_1|^2\right]$. It then follows that one can write $p(x,y)$ in the form

$$p(x,y) = |\det \Omega|^{-1/2} h\left(\left|\begin{pmatrix} \ell_{11}^{-1}\left(x - \omega_{12}\omega_{22}^{-1}y\right) \\ \omega_{22}^{-1/2}y \end{pmatrix}\right|\right)$$

where $\ell_{11} = \omega_{11} - \omega_{12}^2\omega_{22}^{-1}$. Now let

$$\psi(x,y) = x - \omega_{12}\omega_{22}^{-1}y.$$

It is easy to see that the requirement (17) is satisfied. More generally one can take

$$\psi(x,y) = \phi\left(x - \omega_{12}\omega_{22}^{-1}y\right) \tag{18}$$

with $u \mapsto \phi(u)$ is an odd function such that

$$0 < E\left[\phi^2\left(u_1^{\#}\right)\right] < \infty \text{ and } 0 < \kappa = \left.\frac{dE\left[\phi\left(u_1^{\#} + z\right)\right]}{dz}\right|_{z=0} < \infty$$

where

$$u_1^{\#} = u_1 - \omega_{12}\omega_{22}^{-1}\xi_1.$$

Now define

$$\left(\widetilde{\beta}_n^*, \widetilde{\gamma}_n^*\right) = \operatorname{argmin}_{(\beta,\gamma)}\left(Y_k - \beta X_{k-1} - f(\gamma, X_{k-1}) - \omega_{12}\omega_{22}^{-1}\xi_k^*\right)^2$$

with

$$\xi_k^* = \sum_{j=0}^{k-1} d_j(\alpha) v_{k-j}$$

where recall that

$$v_j = X_j - \rho X_{j-1}, \qquad \rho = 1$$

is observable.

With $\left(\widetilde{\beta}_n^*, \widetilde{\gamma}_n^*\right)$ as preliminary estimators, define (similar to $\left(\widehat{\beta}_n, \widehat{\gamma}_n\right)$ defined in the first method) more general M-estimators

$$\begin{pmatrix} \widehat{\beta}_n^* \\ \widehat{\gamma}_n^* \end{pmatrix} = \begin{pmatrix} \widetilde{\beta}_n^* \\ \widetilde{\gamma}_n^* \end{pmatrix} + \left(\kappa \sum_{k=1}^{n} \widetilde{W}_{nk} \widetilde{W}_{nk}^T \right)^{-1} \sum_{k=1}^{n} \widetilde{W}_{nk} \psi \left(Y_k - \widetilde{\beta}_n^* X_{k-1} - f\left(\widetilde{\gamma}_n^*, X_{k-1}\right), \xi_k^* \right). \quad (19)$$

Note that in the case (18)

$$\psi \left(Y_k - \widetilde{\beta}_n^* X_{k-1} - f\left(\widetilde{\gamma}_n^*, X_{k-1}\right), \xi_k^* \right) = \phi \left(Y_k - \widetilde{\beta}_n^* X_{k-1} - f\left(\widetilde{\gamma}_n^*, X_{k-1}\right) - \omega_{12}^2 \omega_{22}^{-1} \xi_k^* \right).$$

One may view the procedure (19) as the one constructed by assuming that the true model is specified by (4) and (5), in which case ξ_k^* are i.i.d.

We first consider the Case III. This means the true model is such that (4) and (5) holds. Then $\xi_k^* \equiv \xi_k, k \geq 1$. The asymptotic behavior of the LS estimators $\widetilde{\beta}_n^*$ and $\widetilde{\gamma}_n^*$ follows exactly as in the Method I of Section 4. (Note that (18) holds with $\phi(x) = x$ in the case of LS estimators.) Explicitly, one has

$$\begin{pmatrix} n^{H+\frac{1}{2}} \left(\widetilde{\beta}_n - \beta \right) \\ \delta_n^{-1} \left(\widetilde{\gamma}_n - \gamma \right) \end{pmatrix} = \begin{pmatrix} \left(n^{-2H-1} \sum_{k=1}^n X_{k-1}^2 \right)^{-1} n^{-H-\frac{1}{2}} \sum_{k=1}^n X_{k-1} u_k^{\#} \\ \left(\delta_n^2 \sum_{k=1}^n f_{(1)}^2(\gamma, a_n X_{k-1}) \right)^{-1} \delta_n \sum_{k=1}^n f_{(1)}(\gamma, a_n X_{k-1}) u_k^{\#} \end{pmatrix} + o_p(1) \quad (20)$$

where

$$u_k^{\#} = u_k - \omega_{12} \omega_{22}^{-1} \xi_k.$$

The limiting distributions of this will have the forms (13) and (14).

Regarding the limiting behavior of $n^{H+\frac{1}{2}}\left(\widetilde{\beta}_n-\beta\right)$ it is important to note that its limit

$$\left(\int_0^1 B_H^2(t)\,dt\right)^{-1}\int_0^1 B_H(t)\,dW(t),$$

where $W(t)$ is now the limit of $n^{-1/2}\sum_{k=1}^{[nt]}u_k^{\#}$, will be mixture of normal distribution. This is because the process $B_H(t)$ will be independent of $W(t)$, or more generally the process $B(t)$ will be independent of $W(t)$ (where recall from the Statement A of Section 2 that $B(t)$ is the limit of the process $n^{-1/2}\sum_{k=1}^{[nt]}\xi_k$) The claimed independence is true because

$$\mathrm{Cov}\left(\xi_1,u_1^{\#}\right)=E\left[\xi_1u_1^{\#}\right]=E\left[\xi_1E\left[u_1^{\#}\middle|\,\xi_1\right]\right]=0.$$

Thus in the present case, in addition to validity of the remark made in Section 3, the both the Wald quadratics for testing the respective parameters β and γ will asymptotically have the central Chi-squared distribution.

The more general M-estimators (19) will have the same asymptotic representation (16) of Section 3 except that the present $\psi(u_k,\xi_k^*)$ will be involved in place of $\psi(u_k)$ of Section 3.

We next consider the Case I. In this case the identity $\xi_k^*\equiv\xi_k, k\geq 1$ does not hold and hence the procedure (19) involves some misspecification. However it can be shown that in this case this misspecification will not have any effect on the asymptotic behavior of the estimators, *under the additional condition* that there is an increasing sequence $k_n\uparrow\infty$ such that

$$\sqrt{n}\sum_{j=k_n}^{\infty}|d_j(\alpha)|\to 0.$$

Actually when this condition holds ξ_k^* involved in the construction (19) can be replaced by $\xi_k^{**}=\sum_{j=0}^{\min(k-1,k_n)}d_j(\alpha)v_{k-j}$.

Thus the conclusions of the preceding Case III remain true for this case.

We now consider the Case II: In this case also,the procedure (19) involves the misspecification indicated in the preceding case I. But unfortunately it introduces a sort of bias which cannot be eliminated as in the Case I.

To indicate the nature of the bias let

$$R_k = \sum_{j=k}^{\infty} d_j(\alpha) v_{k-j}.$$

Note that this quantity does not depend on the parameters β and γ. One has

$$\xi_k = \xi_k^* + R_k$$

so that

$$u_k^* = u_k - \omega_{12}\omega_{22}^{-1}\xi_k^* = u_k^{\#} + \omega_{12}^2\omega_{22}^{-1}R_k.$$

where recall that $u_k^{\#} = u_k - \omega_{12}\omega_{22}^{-1}\xi_k$. Let us first consider the LS estimator $\widetilde{\beta}_n^*$. Suppose for the moment that the representation

$$n^{H+\frac{1}{2}}\left(\widetilde{\beta}_n^* - \beta\right) = \left(n^{-2H-1}\sum_{k=1}^{n} X_{k-1}^2\right)^{-1} n^{-H-\frac{1}{2}}\sum_{k=1}^{n} X_{k-1}u_k^* + o_p\left(1\right)$$

is true. Then note that

$$n^{-H-\frac{1}{2}}\sum_{k=1}^{n} X_{k-1}u_k^* = n^{-H-\frac{1}{2}}\sum_{k=1}^{n} X_{k-1}u_k^{\#} + \omega_{12}^2\omega_{22}^{-1}n^{-H-\frac{1}{2}}\sum_{k=1}^{n} X_{k-1}R_k.$$

It can be shown that the limit will be of the form

$$\int_0^1 B_H\left(t\right)dW\left(t\right) + A$$

with $W\left(t\right)$ independent of $\left(B_H\left(t\right), A\right)$. Thus the limiting distribution of $n^{H+\frac{1}{2}}\left(\widetilde{\beta}_n^* - \beta\right)$ will be a mixture of non-central Chi-squared distribution (with a random noncentrality parameter). We have not yet obtained an explicit expression for the limit R. Similarly, regarding the LS estimator $\widetilde{\gamma}_n^*$ there will be a bias term of the form

$$\delta_n \sum_{k=1}^{n} f_{(1)}(\gamma, a_n X_{k-1})R_k$$

At the moment we have no idea of how to handle this, unfortunately, except to show that

$$\sup_n E\left[\left|\delta_n \sum_{k=1}^{n} f_{(1)}(\gamma, a_n X_{k-1}) R_k\right|\right] < \infty.$$

(We do not believe that $\delta_n \sum_{k=1}^{n} f_{(1)}(\gamma, a_n X_{k-1}) R_k$ will converge to zero in probability.) Thus the procedure (19) does not seem to be satisfactory for the Case II.

5 LAMN STRUCTURE OF THE LIKELIHOOD

Let $P_{\theta,n}$ be the joint distribution of the observables $((X_k, Y_k), k = 1, \ldots, n)$ where we let $\theta = (\beta, \gamma)'$. Define the log-likelihood ratios:

$$\Lambda_n(\theta^*, \theta) = \log \frac{dP_{\theta^*,n}}{dP_{\theta,n}}.$$

Also let

$$\delta_{*n} = \text{diag}\left\{n^{H+\frac{1}{2}}, \delta_n\right\}.$$

Further let

$$\psi(x, y) = -\frac{\partial \log p(x, y)}{\partial x}$$

and

$$0 < \lambda = E\left[\left|\psi(\eta_1, \xi_1)\right|^2\right] < \infty$$

where recall that the $p(x, y)$ is joint density function of (u_1, ξ_1).

One has

Statement C: For any bounded 2-vectors $h_n = (h_{n1}, h_{n2})$, one has the following conclusions:

(a)

$$\begin{aligned}
\Lambda_n(\theta + \delta_{*n} h_n, \theta) &= h_n' \delta_{*n} W_n^{\#}(\theta) - \frac{1}{2} h_n' \delta_{*n} M_n(\theta) \delta_{*n} h_n + o_p(1) \\
&= h_{n1} n^{-H-\frac{1}{2}} \sum_{k=1}^{n} X_{k-1} \psi(u_k, \xi_k) - \frac{\lambda}{2} h_{n1}^2 n^{-2H-1} \sum_{k=1}^{n} X_{k-1}^2 \\
&\quad + h_{n2} \delta_n \sum_{k=1}^{n} f_{(1)}(\gamma, a_n X_{k-1}) \psi(u_k, \xi_k) - \frac{\lambda}{2} h_{n2}^2 \delta_n^2 \sum_{k=1}^{n} f_{(1)}^2(\gamma, a_n X_{k-1}),
\end{aligned}$$

where

$$\begin{aligned} W_n^{\#}(\theta) &= \sum_{k=1}^{n} \widetilde{W}_{nk}\psi\left(u_k, \xi_k\right) \\ &= \sum_{k=1}^{n} \begin{pmatrix} X_{k-1}\psi\left(u_k, \xi_k\right) \\ f_{(1)}(\gamma, a_n X_{k-1})\psi\left(u_k, \xi_k\right) \end{pmatrix} \end{aligned}$$

(recall that $u_k = Y_k - \beta X_{k-1} - f\left(\gamma, X_{k-1}\right)$), and

$$M_n(\theta) = \lambda \sum_{k=1}^{n} \widetilde{W}_{nk}\widetilde{W}_{nk}^T.$$

Note that $W_n^{\#}(\theta)$ involves variables ξ_k that are not completely observable.

(b) $\delta_{*n}W_n^{\#}(\theta)$ and $\delta_{*n}M_n(\theta)\delta_{*n}$ jointly converge in distribution to the limits (13) and (14) with their meanings as described in Section 4.

(c) The sequences $\{P_{\theta,n}\}$ and $\{P_{\theta+\delta_{*n}h_n,n}\}$ are contiguous.

(d)

$$\delta_{*n}W_n^{\#}(\theta + \delta_{*n}h_n) = \delta_{*n}W_n^{\#}(\theta) - \delta_{*n}M_n(\theta)\delta_{*n}h_n + o_p(1)$$

and

$$\delta_{*n}M_n(\theta + \delta_{*n}h_n)\delta_{*n} = \delta_{*n}M_n(\theta)\delta_{*n} + o_p(1)$$

in $P_{\theta,n}$.

(e) In the Cases I and III of Section 2

$$\delta_{*n}W_n^{\#}(\theta) = \delta_{*n}\sum_{k=1}^{n} \widetilde{W}_{nk}\psi\left(Y_k - \beta X_{k-1} - f\left(\widetilde{\gamma}, X_{k-1}\right), \xi_k^*\right) + o_p(1)$$

in $P_{\theta,n}$, where recall that $\xi_k^* = \sum_{j=0}^{k-1} d_j(\alpha)v_{k-j}$ which is observable whereas ξ_k is not completely observable.